Research Note 84-87

AD-A142 434

ANALYSIS OF COMPUTER INTERACTIVE TESTS FOR ASSIGNING HELICOPTER PILOTS TO DIFFERENT MISSIONS

David C. Myers, F. Mark Schemmer, and Edwin A. Fleishman
Advanced Research Resources Organization

Michael G. Sanders, Contracting Officer's Representative

Submitted by

Charles A. Gainer, Chief
ARI FIELD UNIT AT FORT RUCKER, ALABAMA

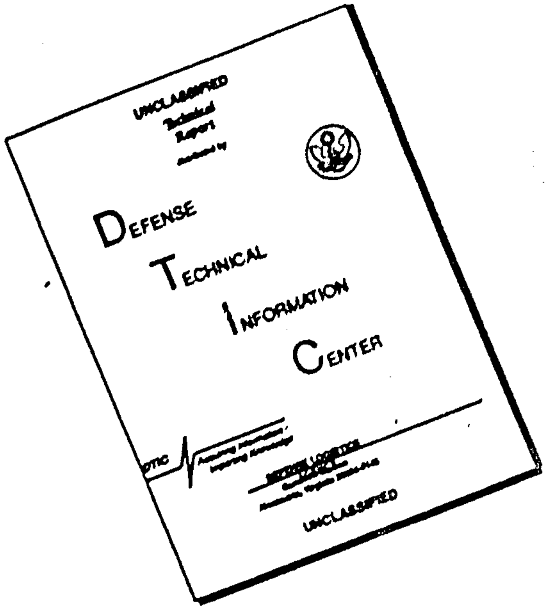
and

Harold F. O'Neil, Jr., Director
TRAINING RESEARCH LABORATORY

U. S. Army
Research Institute for the Behavioral and Social Sciences

June 1984

DISCLAIMER NOTICE


UNCLASSIFIED
Technical
Report

DEFENSE
TECHNICAL
INFORMATION
CENTER

DTIC

UNCLASSIFIED

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Research Note 84-87	2. GOVT ACCESSION NO.: AD-A142434
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Analysis of Computer Interactive Tests for Assigning Helicopter Pilots to Different Missions	5. TYPE OF REPORT & PERIOD COVERED: Research Note 11/2/81 - 5/2/83
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): David C. Myers F. Mark Schemmer Edwin A. Fleishman	8. CONTRACT OR GRANT NUMBER(s): MDA903-82-C-0036
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Advanced Research Resources Organization 4330 East West Highway, Suite 900 Bethesda, Maryland 20814	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q263731A792
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Supply Service-Washington Room 1D245, The Pentagon Washington, DC 20310	12. REPORT DATE: June 1984
	13. NUMBER OF PAGES: 46
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Army Research Institute Field Unit ATTN: Dr. M. Sanders (PERI-IR) Ft. Rucker, AL 36362	15. SECURITY CLASS. (of this report): N/A
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

N/A

18. SUPPLEMENTARY NOTES

N/A

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

[illegible]ities Analysis
[illegible]ssion Track Assignment Battery
[illegible]ter-Interactive Tests
[illegible]aring
[illegible]
Pilot Performance Measures
Reliability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

[illegible] Reverse Side

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A test battery was developed to represent a broad range of abilities and skills that were identified as important in piloting helicopters. The battery was developed on the basis of a taxonomic approach to job analysis which linked pilot tasks to ability requirements for different mission tracks. The tests developed were based on an earlier review, by ARRO staff, of the kinds of tests likely to measure the abilities identified in the job analysis approach employed. The present study included the development, programming and pre-testing of computer interactive tests designed to measure abilities identified as underlying critical tasks in the various helicopter missions. Optimum conditions of administration were developed and test reliability was determined. Tasks judged by expert pilots to be critical for pilot effectiveness were identified as possible measures of performance in the different mission tracks. These tasks can be translated into criterion measures and used in validating the test battery developed. The purpose of the validation will be to empirically link test scores with performance in these critical tasks representing the different mission tracks. The findings would indicate if all the tests are needed or if empirical validities showed which limited set of tests can be used to predict mission performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The original objectives of the research included: (1) development of performance-based measures for assessing mission specific skills, knowledges, and abilities of student pilots in the tactics phase of the Initial Entry Rotary Wing training program; (2) development of performance-based measures for assessing mission proficiency of mission track training program graduates shortly after arrival at their first unit assignment; and (3) determination of the predictive validity of the Mission Track Assignment Battery for each of the four missions in terms of institutional (i.e., training) and operational (i.e., field assignment) criteria. During the early phases of the project it became apparent that it was necessary to modify these original objectives. The actual nature of the tasks and the level of abilities required to pilot helicopters in the different missions were not clear; and the existing test battery had not been fully developed and had not been linked to the ability requirements. Therefore, it was decided to spend more time and effort to identify the critical pilot tasks and ability requirements, and to develop and adapt the test battery to computer technology. It was believed that this new research plan would insure that the pilot proficiency evaluations would, when performed, be based on thoroughly researched and constructed tests and criteria. Those resources which would have been used to collect the criterion data in the field and in training were used to identify critical tasks and ability requirements, as well as to develop a computer-based test battery.

The data collection for this project was carried out at Fort Rucker, Alabama, with the cooperation of officials in the IERW training program. The Army Research Institute Contracting Officer Technical Representative on the project was Dr. Michael G. Sanders. He and Dr. Jack A. Dohme, of ARI provided valuable assistance in all phases of the research effort. We would like also to acknowledge the assistance of Ms. Margarette Jennings of ARRO who helped in the job analysis, Mr. Thomas Folks who helped design some of the computer software, and Mrs. Jackie Allums and Mrs. Dorothy Churchman who assisted in the test administration at Fort Rucker.

SUMMARY

A test battery was developed to represent a broad range of abilities and skills that were identified as important in piloting helicopters. The battery was developed on the basis of a taxonomic approach to job analysis which linked pilot tasks to ability requirements for different mission tracks. The tests developed were based on an earlier review, by ARRO staff, of the kinds of tests likely to measure the abilities identified in the job analysis approach employed. The present study included the development, programming and pre-testing of computer interactive tests designed to measure abilities identified as underlying critical tasks in the various helicopter missions. Optimum conditions of administration were developed and test reliability was determined. Tasks judged by expert pilots to be critical for pilot effectiveness were identified as possible measures of performance in the different mission tracks. These tasks can be translated into criterion measures and used in validating the test battery developed. The purpose of the validation will be to empirically link test scores with performance in these critical tasks representing the different mission tracks. The findings would indicate if all the tests are needed or if empirical validities showed which limited set of tests can be used to predict mission performance.

INTRODUCTION

The importance of rotary wing aircraft to the accomplishment of Army missions has been increasing. In the past, helicopters served a secondary role in battlefield operations such as support for ground forces. More recently, helicopter operations have been expanded to include destruction of mobile and small targets, particularly tanks. Pilots are also required to fly at altitudes several meters above ground, at night and during the day as well as carry out missions for extended periods of time (Stone, Kruger, & Hold, 1982). Because of the expanded role being carried out by helicopter pilots, there has also been an increasing diversification associated with achieving specific mission objectives. Although most helicopter pilots are required to perform similar flight tasks such as instrument takeoff, masking, and NOE flight, there are specialized tasks performed only by those pilots assigned to a particular mission. For example, while Aeroscout pilots perform aerial and zone reconnaissance, Attack pilots clear weapon systems and operate TOW missile.

Because of the increasingly specialized role played by helicopter pilots in their various missions there is a need for improved procedures to assign student pilots to the missions in which they will be most effective. The purpose of the project described in this report was to develop a battery of tests that would allow the Army to assign pilots to one of four missions--Aeroscout, Attack, Cargo, and Utility--depending on the match between mission requirements and the skills of the individual.

Piloting a helicopter is a complex and demanding job. The inherent instability of the helicopter requires almost continuous use of the controls, which interact intricately with each other. In virtually all operations the three basic controls must be operated simultaneously. At the same time, instrument readings must be made and specific positions with respect to the ground must be attained. Therefore, because of the complexity of the aircraft flown and the changing requirements of the pilot's job, there is a need to better understand skill components of

helicopter proficiency (Zavala, Locke, Van Cott, & Fleishman, 1965; Locke, Zavala, & Fleishman, 1965). This could be accomplished by applying more sophisticated methods in the analysis of their skill, the specification of the abilities required, and the linkage of test methods to evaluate these abilities.

The current Initial Entry Rotary Wing (IERW) training program at the US Army Aviation Center (USAAVNC) is a dual track course in which 25% of student pilots (SPs) graduate as Aeroscout aviators and the remainder graduate as Utility aviators. The dual track program has been evaluated in a survey-based research effort and was found to be a cost-effective training technique. Thus, USAAVNC tasked Army Research Institute (ARI) to develop a means of testing and assigning SPs to a Mission Track IERW training program in which aviators would earn their wings in one of four helicopter missions: Aeroscout, Attack, Cargo or Utility (Figure 1).

Objectives

To meet this need the present research project was undertaken to develop procedures for testing and assigning student pilots to specific tracks in the Initial Entry Rotary Wing (IERW) training program. The Mission Track Assignment Battery (MTAB) was developed to classify student pilots to one of four missions. The following were the objectives of the research effort.

- Determination of critical tasks performed in the different missions.
- Identification of ability requirements of piloting helicopter at the mission level.
- Development of a battery of tests, called the MTAB, designed to assess skills, and abilities required of pilots.
- Development and utilization of a computer-interactive testing system.
- Tryout of the Mission Track Assignment Battery.
- Development of recommendations with regard to testing procedures, scoring, and administration.

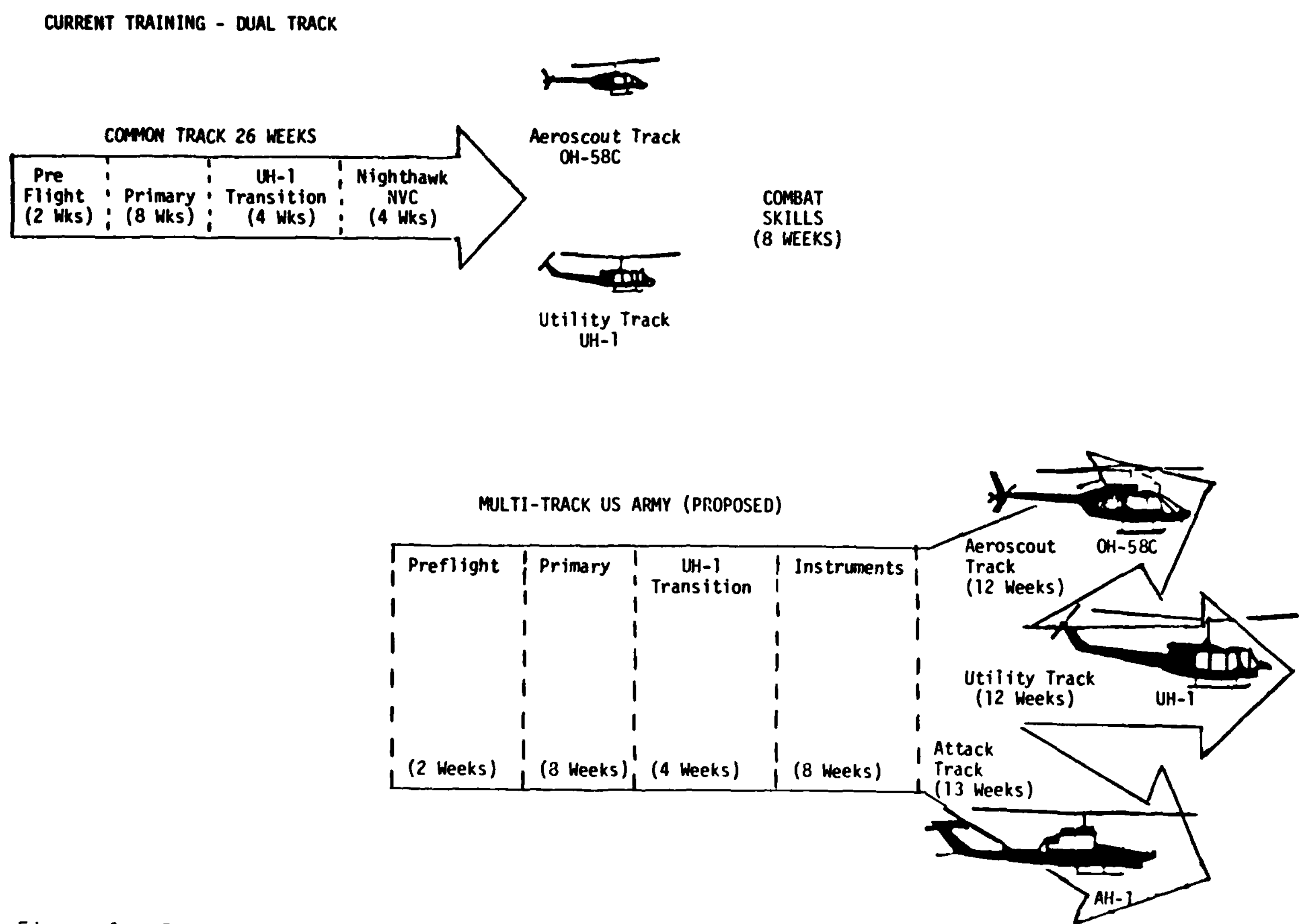


Figure 1. Proposed multi-track training program.

Background

Over the years, psychologists have addressed methodological issues and researched operational problem areas associated with the selection and training of aviators in the military. The selection procedures developed predict up to 25 percent of the variance in pilot performance at advanced stages in training (Roscoe & North, 1980). With the increasing technical sophistication of aircraft and the apparent problems with pilot combat effectiveness (Youngling, Levine, Mocharnuk, & Weston, 1977), there is increasing consideration being given to improving the procedures used to select and assign pilots to training missions base on a more comprehensive evaluation of their skills and abilities. S effort would improve the military's use of manpower and help to enha the safety and effectiveness of their aviation program.

Although there has been some success in predicting pilot performance and combat effectiveness, a number of hypotheses reporting additional relevant skills have been proposed. Some "desirable traits" mentioned include the ability to deal with emergencies while not losing control of routine tasks; ability to estimate quickly probable outcomes for different courses of action; ability to reorder priorities as situations deteriorate or improve; ability to take decisive action in the face of indecision by others. The problem is to evaluate through research which of a wide range of potentially useful abilities are relevant to pilot success.

There is considerable research that suggests perceptual-motor measures may contribute significantly to the prediction of pilot proficiency. In addition, the development of electronic components and the increasing availability of low-cost computer terminals have eliminated the difficulties formerly associated with the assessment of perceptual-motor abilities. For example, McGrevy and Valentine (1974) adapted the Two-Hand Coordination and Complex Coordination tests employed by Melton (1947) and Fleishman (1956) to such a device. Performance on these tests was found correlated with a variety of flight criteria. Melton (1947) has summarized research on the various tests developed by the

Army, and the Air Force in World War II for selection of various air crew personnel. This work showed a number of perceptual-motor tests to be valid for predicting success in flight-training school for pilots, including standard classification tasks such as Complex Coordination (CM 201), Two-Hand Coordination (CM 101B), Discrimination Reaction Time (CP 611D), Rotary Pursuit (CM 803B), and Rudder Control (CM 120B).

In the post-World War II period of development of psychological tests for pilot selection Fleishman (1956) employed a factor-analytic approach to identify the ability factors which underly psychomotor test performance. He also examined the possibilities of using printed tests to duplicate such variance found in apparatus tests. Tests most diagnostic of the different perceptual-motor abilities identified were developed.

Fleishman and his associates demonstrated validity for a number of new perceptual-motor apparatus tests (e.g., Fleishman, 1954). Moreover, they showed that the validity of these tests does not require "job sample" type tests. Rather tests of the underlying perceptual-motor and cognitive abilities underlying pilot performance allows construction of tests to measure the relevant abilities more precisely, and hence to select pilots more efficiently (Fleishman, 1956). For example, they obtained scores made by student pilots on 24 standard maneuvers, and a factor analysis of these scores revealed six factors already identified common to tests and criterion pilot performance, i.e., control precision, multi-limb coordination, rate control, spatial relations, response orientation, and procedural integration (Fleishman & Ornstein, 1960). A later, more comprehensive analysis of this data also revealed a kinesthetic discrimination factor along with other piloting factors (Zavala, Locke, Van Cott, & Fleishman, 1965).

Contemporary aircraft have become increasingly sophisticated with technological advances, automatizing many functions previously performed by the pilot, but also increasing the load placed on the pilot's cognitive functioning. The speed and accuracy with which information is

perceived, encoded, stored, transformed, and compared, the speed with which memory is searched and accessed, and the speed with which appropriate decisions may be made, are all crucial to pilot performance. Thus, cognitive abilities would appear worthy of consideration for predicting helicopter pilot proficiency. In a study recently conducted at ARRO for the Air Force a number of cognitive processes were identified as important to successful piloting (e.g., attention and decision-making). The same report included recommendation for development of a battery of tests designed to assess these processes (Imhoff & Levine, 1980).

Recognition of the importance of cognitive abilities is reflected in the literature on pilot training and selection. For example, current pilot training research places a great deal of emphasis on cognitive pre-training techniques to improve comprehension and integration of necessary aviation information (Gerlach, 1974; Crosby, 1977), and programs for judgment training have been formulated (Jensen & Benel, 1977). Selection tests have focused on selective attention (Gopher & Kahneman, 1971) and timesharing tasks (North & Gopher, 1976) as predictors of pilot performance, but recent test batteries have included a number of tasks involving memory, spatial visualization, comprehension, and other cognitive functions (McLaurin, 1973; Pew & Adams, 1975; Hunter, 1975). Workload assessment has also been given increasing consideration by a number of investigators (Damos, 1978; Wierwille & Connor, 1983; Levine, Ogden, & Eisner, 1978).

The present project dealt with these issues in developing tests potentially useful in predicting helicopter performance. The study also draws on the job analysis methods developed earlier (see Fleishman, 1975,1982; Fleishman, Quaintance, & Broedling, 1983; Theologous, Romashko, & Fleishman, 1973) for converting information about job tasks into ability requirements and test procedures. This approach will be described in the next section.

Project Phases

The research project was conducted in three phases. In Phase one job analytic efforts to determine the critical tasks and relevant abilities for each of the four mission tracks were carried out. The second phase entailed the development of the ten tests in the battery as well as the design of the apparatus and testing procedures. In Phase three efforts were made to evaluate the MTAB in an operational setting involving testing of over 300 student pilots prior to IERW training. This phase also included analysis of the results from the large scale administration. Two interim reports describe phases one and two in detail (Myers, Jennings, & Fleishman, 1982; Myers, Jennings, Schemmer, & Fleishman, 1982). Therefore, the present final report summarizes only the methodology and findings for these first two phases of the project while focusing more on the third phase which involved analysis of the results from the administration of the MTAB.

PHASE I. JOB ANALYSIS

This section of the report presents the results of our efforts to identify the "important" tasks and the abilities considered necessary to successfully perform these tasks for each of the mission tracks. It includes the procedures used to obtain the task information and the methods used to analyze the ability analysis data. The section concludes with a presentation of the most important abilities.

Identification of Critical Tasks

Based upon a review of the work accomplished by Miller, Eschenbrenner, Marco, and Dohme (1981), it was concluded that although the research provided some useful data regarding task criticality, the authors failed to include a broad comprehensive range of task statements covering all aspects of the pilot's mission. Consequently, it was necessary to begin the present project with a task analysis based on a more comprehensive list of tasks.

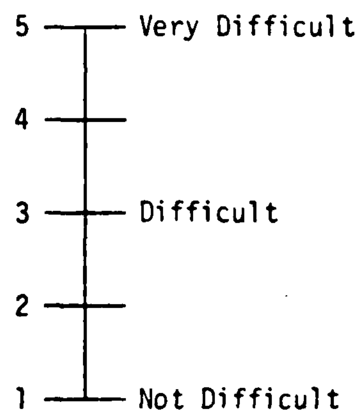
The training documents and materials such as Aircrew Training Manuals (ATM), Army Training and Evaluation Program (ARTEP), and Operator's Manuals were reviewed to gain an understanding of helicopter operations, training procedures, and Army regulations. The ATMs were especially useful for describing the job tasks required of the pilots performing the four different missions. Each ATM consisted of a series of tasks including basic flight tasks, emergency procedures, instrument flight and mission tasks. All of the tasks incorporated in the mission and special/tactical section of the ATMs were used to develop a task survey instrument for each of the four missions (i.e., Attack TC1-136; Utility TC1-138; Cargo TC1-139; and Observation TC1-137). The tasks in the ARTEPs were not used because they appeared to be covered by the ATM task lists.

Each survey was designed to obtain the judgments of expert pilots regarding two task variables--Difficulty and Consequence of Inadequate Performance (see Figures 2 and 3). On February 9, 1982, a two-hour meeting was convened to assess task criticality. The participants

This scale measures task difficulty. The difficulty of tasks depend upon such features as the degree of repetitiveness, variation of problems encountered, as well as the range of skills utilized and subject matter knowledge drawn upon. The ratings should be made based on your own abilities and perceptions about task difficulty. Although all tasks in the list may be considered difficult, we are interested in the relative difficulty between the tasks. The task is to be rated on a scale from "1" (Not Difficult) to "5" (Very Difficult).

How difficult is the task?

Task involves constantly dealing with new problems and situations which require a wide range of skills.



Task involves dealing with routine situations which require relatively few skills.

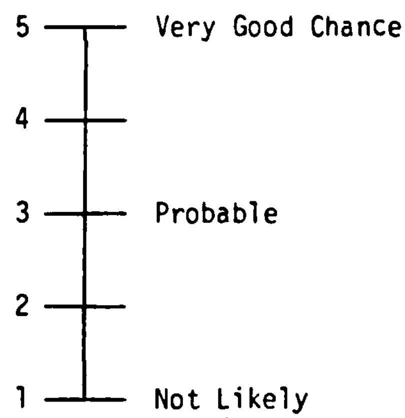
PLEASE TURN THE PAGE AND RATE EACH OF THE FOLLOWING TASKS USING THE 5-POINT SCALE SHOWN ABOVE.

Figure 2. Scale for evaluating task difficulty.

This scale is a measure of the probability of serious consequences of inadequate performance of a task. It is related to the importance of the task for the accomplishment of the pilot's mission. It is defined in terms of the chances that inadequate performance will lead to mission failure, injury or death, wasted supplies and/or damaged equipment. The task is to be rated on a scale from "1" (serious consequences are Not Likely) to "5" (Very Good Chance for serious consequences).

What will happen if the task is inadequately performed?

Inadequate performance of the task will most likely lead to serious consequences (failure to complete the mission, injury, death, damaged equipment).



Inadequate performance of the task will probably not lead to serious consequences (failure to complete the mission, injury, death, damaged equipment).

PLEASE TURN THE PAGE AND RATE EACH OF THE FOLLOWING TASKS USING THE 5-POINT SCALE SHOWN ABOVE.

Figure 3. Scale for evaluating consequences of inadequate performance.

consisted of Warrant Officers attending refresher training at Ft. Rucker. Under the guidance of ARRO staff, four groups of expert pilots who represented the different missions rated the tasks taken from the ATMs (Attack N = 15; Aeroscout N = 12; Cargo N = 17; Utility N = 20). During the data analysis a decision rule was developed to determine the most critical tasks for each mission. For each task, the two mean ratings (i.e., difficulty and consequences of inadequate performance) were averaged to obtain an overall index of task importance because the task variables were significantly correlated (i.e., Attack r = .41; Utility r = .73; Cargo r = .69; and Observation r = .59).

The tasks representing each mission were ranked according to these overall ratings of importance. The results of the task analysis are described in a previous report (Myers et al. 1982). The most important tasks were those ranked highest in each misson track (e.g., top five in the tactical/special group). These tasks were used to determine the ability requirements of each mission, and were made available for future development of criterion measures which represent pilot proficiency.

Determination of Ability Requirements

In addition to determining ability requirements based upon existing literature, a methodology developed at ARRO, and supported by considerable research evidence, was used to systematically link ability requirements to the critical pilot tasks. ARRO staff have developed procedures to translate task characteristics into ability taxonomies for predicting human performance (Fleishman, 1975, 1982). In earlier research, 11 psychomotor and 9 physical abilities were identified as accounting for a substantial amount of variance in performance in a wide range of tasks. These abilities were identified on the basis of the correlations among performance in an extensive series of studies involving more than 200 different tasks (Fleishman 1964, 1973). In subsequent studies the definitions of these abilities and their distinctions from one another have been more clearly delineated. Abilities in the cognitive and perceptual domain have also been identified (Theologus, Romashko, & Fleishman, 1973).

Subsequent research (Fleishman & Hogan, 1978; Myers, Gebhardt, & Fleishman, 1979) has developed techniques for determining the extent to which such abilities are required for performance on complex jobs. A series of rating scales contained in a manual has been developed for converting information about jobs into these requirements. Each scale consists of a definition of the ability, a comparison with other abilities, and a 7-point scale where "1" represents low levels of the ability and "7" represents high levels of the ability. The task examples and their location on the scales have been empirically determined by previous research. Figure 4 illustrates the type of scale used to determine the task requirements.

For the present study, a meeting with 57 expert pilots (i.e., Aeroscout = 13; Attack = 16; Cargo = 11; and Utility = 17) was convened on March 12, 1982. Under the guidance of research staff, the pilots rated each task deemed critical to their mission in terms of its ability requirements (e.g., divided attention, control precision, and deductive reasoning). Each ability scale had seven points and included several task anchors to assist the rater. The data analysis indicated the abilities receiving the highest ratings within each mission, as well as those abilities which best differentiated between the four missions. Table 1 provides a comparison of the perceived ability requirements between the different mission tracks. These means were determined by averaging the ratings across all of the critical tasks representing a particular mission. The results indicated that the abilities required to pilot helicopters may vary between missions. Perceptual Speed, for example, received a higher rating for the Attack mission than the other missions. Multilimb Coordination and Control Precision were important for Cargo and Utility missions. Divided Attention was one of the most important abilities for the Aeroscout mission.

This is the ability to adjust an equipment control in response to changes in the speed and/or direction of a continuously moving object or scene. The ability involves timing these adjustments in anticipating these changes.

This ability does not extend to situations in which both the speed and direction of the object are perfectly predictable.

How Rate Control Is Different from Other Abilities

Involves timing of continuous movements.	vs.	Control Precision (15): Involves quick adjustments of equipment controls to exact positions.

Requires precisely timed, fine motor adjustments to random changes of a high speed object moving in several directions.

7

--Operate aircraft controls to land a jet on aircraft carrier in turbulent weather.

6

5

--While hunting, shoot a duck in flight.

4

--Keep up with a car you are following where the speed of the first car may vary.

3

--Ride a bicycle alongside a runner.

2

1

Requires timed motor adjustments to a slow moving, almost predictable object moving in a single direction.

Figure 4. Rating scale for rate control.

TABLE 1

Summary of Mean Ability Ratings by Mission

	Ability Name	Ability Code	Mission Attack	Mission Cargo	Mission Observ.	Mission Utility
1.	Written Expression	WE	2.55	2.49	2.30	2.01
2.	Written Comprehension	WC	3.04	3.03	2.55	2.46
3.	Oral Expression	OE	3.59	3.02	2.94	3.16
4.	Oral Comprehension	OC	3.57	3.38	2.66	3.25
5.	Perceptual Speed	PSP	4.02	2.86	3.43	3.54
6.	Visualization	V	3.54	3.43	3.15	3.70
7.	Spatial Orientation	SO	3.39	3.72	3.63	3.70
8.	Divided Attention	DA	4.11	3.75	3.88	4.04
9.	Selective Attention	SA	3.77	3.54	3.79	3.71
10.	Flexibility of Closure	FC	3.76	3.17	3.57	3.66
11.	Speed of Closure	SC	3.58	3.60	3.54	3.70
12.	Reaction Time	RT	3.80	3.59	3.15	4.04
13.	Choice Reaction Time	CR	3.83	3.68	3.20	3.73
14.	Multilimb Coordination	ML	3.48	4.01	3.20	4.25
15.	Control Precision	CP	3.52	4.03	3.16	3.99
16.	Kinesthetic Memory	KM	3.47	3.67	3.10	4.27
17.	Rate Control	RC	3.27	3.66	3.23	3.99
18.	Arm-Hand Steadiness	AH	2.93	3.59	2.82	3.46
19.	Finger Dexterity	FD	2.86	2.96	2.13	2.68
20.	Speed of Limb Movement	SL	2.45	2.76	2.66	3.56
21.	Wrist-Finger Speed	WF	1.38	2.29	2.51	3.46
22.	Memorization	M	3.43	3.28	3.56	3.83
23.	Decision Making	DM	3.69	3.90	3.54	4.26
24.	Information Ordering	IO	3.13	3.11	3.05	3.66
25.	Category Flexibility	CF	3.08	2.64	3.13	3.54
26.	Number Facility	NF	2.70	2.89	2.43	2.81
27.	Problem Sensitivity	PS	3.30	3.89	3.26	3.91
28.	Deductive Reasoning	DR	3.40	3.80	3.24	3.67
29.	Inductive Reasoning	IR	3.14	3.45	2.89	3.51
30.	Originality	OR	2.86	2.93	2.82	3.13
31.	Fluency of Ideas	FI	2.99	2.86	2.97	3.52
32.	Stress Tolerance	ST	4.10	4.40	3.94	4.08

Note: Means represent average of ratings across all of the important tasks.

PHASE II. TEST DEVELOPMENT

Abilities Differentiating Missions

The first step in the test development process involved a discriminant function analysis which determined the abilities best differentiated between the four missions (i.e., stepwise analysis, BMDP7M). A seven-step solution provided the best interpretation of the data. Table 2 presents the summary statistics derived from the discriminant analysis.

Since a major objective of the data analysis was to identify a limited number of the most important abilities for test development, the mean ratings were again averaged across the different missions allowing for an overall ranking of the abilities. Table 3 includes the results of this ranking and illustrates the abilities which discriminated significantly. It also indicates the abilities which were to be translated into tests using the computer mode and/or paper-pencil format. Some caution is required in interpreting these findings because of the sample sizes in the sub-groups.

The abilities considered most important for test development were selected first by including those which discriminated and then by including those abilities which were rated highest based on averages across the different missions. Thus, Position Memory, Spatial Orientation, Perceptual Speed, and Flexibility of Closure were selected because they received higher rankings and were significant discriminators. The three remaining discriminators were excluded because they had received low overall ratings, and measures designed to assess these abilities would probably not correlate highly with criterion measures of pilot performance. Since the remaining abilities did not significantly differentiate between missions, they were selected on the basis of magnitude of their average ratings (Table 3). It should be noted that the present state-of-the-art did not provide reliable tests which were ready for immediate implementation to measure stress tolerance and decision making; therefore, these factors were eliminated from further analysis.

TABLE 2

Summary Table for Discriminant Function Analysis

Variable Entered	F-Value To Enter Or Remove	Number of Variables Included	U-Statistic	Approximate F-Statistic	Degrees of Freedom	
Wrist Finger Speed	2.96	1	.85	2.96	3.00	50.00
Written Expression	2.75	2	.73	2.83	6.00	98.00
Oral Comprehension	2.22	3	.64	2.64	9.00	116.97
Spatial Orientation	2.68	4	.54	2.68	12.00	124.64
Flexibility of Closure	3.03	5	.46	2.80	15.00	127.39
Kinesthetic Memory	2.92	6	.38	2.89	18.00	127.76
Perceptual Speed	2.24	7	.33	2.84	21.00	126.89

TABLE 3

Ranking of Abilities Based on Averages Across the Different Missions

ABILITY	ACROSS MISSIONS MEANS	ACROSS MISSIONS STANDARD DEVIATION	RANK-ORDERED ABILITIES WHICH SIGNIFICANTLY DIFFERENTIATED BETWEEN MISSIONS[b]	ABILITIES RECOMMENDED FOR TEST DEVELOPMENT (INCLUDING DISCRIMINATORS)
STRESS TOLERANCE	4.12[a]	1.05		-
DIVIDED ATTENTION	3.97	1.13		*
DECISION MAKING	3.86	1.04		-
MULTILIMB COORDINATION	3.75	1.08		*
SELECTIVE ATTENTION	3.71	1.02		*
REACTION TIME	3.67	1.02		*
KINESTHETIC MEMORY	3.66	1.12	6	*
PERCEPTUAL SPEED	3.66	.97	7	*
CONTROL PRECISION	3.65	1.08		*
CHOICE REACTION TIME	3.62	1.00		*
FLEXIBILITY OF CLOSURE	3.60	1.02	5	*
PROBLEM SENSITIVITY	3.60	1.15		*
SPATIAL ORIENTATION	3.59	1.07	4	*
SPEED OF CLOSURE	3.59	1.09		*
MEMORIZATION	3.59	1.19		*
DEDUCTIVE REASONING	3.52	1.00		
RATE CONTROL	3.52	.94		
VISUALIZATION	3.47	1.11		
INFORMATION ORDERING	3.25	1.06		
INDUCTIVE REASONING	3.24	1.04		
ORAL EXPRESSION	3.23	1.07		
ORAL COMPREHENSION	3.22	1.07	3	
CATEGORY FLEXIBILITY	3.19	1.12		
ARM-HAND STEADINESS	3.17	1.13		
FLUENCY OF IDEAS	3.13	1.21		
ORIGINALITY	2.91	1.17		
SPEED OF LIMB MOVEMENT	2.85	1.20		
WRIST-FINGER SPEED	2.77	1.22	1	
WRITTEN COMPREHENSION	2.75	1.22		
MEMBER FACILITY	2.69	1.40		
FINGER DEXTERITY	2.61	1.19		
WRITTEN EXPRESSION	2.38	1.40	2	

NOTE: [a]MEAN REPRESENTS AVERAGE ACROSS FOUR MISSIONS

[b]BASED ON DISCRIMINANT FUNCTION ANALYSIS (ORDER OF INCLUSION 1 THROUGH 7)

Interrater Reliability

Estimates of the amount of agreement among the expert raters who provided the ability by task ratings were used as an additional basis for developing the test battery. The intraclass correlations were computed to estimate the reliability of the mean task ratings. The specific computational method used to estimate the reliabilities corresponds to the ICC(2,1) coefficient presented in Shrout and Fleiss (1979). This particular estimation assumes random effects for raters and for tasks. Any variance which is a function of between-rater mean differences or rater-by-task interaction is considered error variance, while between-task mean variance is assumed to reflect true variance.

The coefficients resulting from this procedure were correlation coefficients representing the estimated reliabilities of any single task rating by an individual rater. However, the model relied upon mean task ratings, and it is well known that mean ratings are more reliable than single ratings. The estimated reliability of a mean is appropriately obtained by applying the Spearman-Brown correlation formula to the estimated reliability of individual ratings (cf. Winer, 1971; Shrout & Fleiss, 1979). In order that the resulting coefficients would be comparably scaled, a single number of raters were consistently applied in making this transformation. An examination of the number of experts rating each task indicated that very few tasks were rated by less than ten experts. Consequently, an N of 10 was used in the Spearman-Brown transformations. Of course, this transformation is positive and monotonic with respect to N, so that mean task ratings arising from larger Ns are, in general, even more reliable than the estimates gained in this manner.

The estimated reliability coefficients presented in Table 4 indicate substantial agreement among the pilots who provided the independent ability ratings. There were some exceptions where scales had low reliabilities such as written/oral comprehension and expression. These abilities were not recommended for development of tests. It should be

TABLE 4

Interrater Reliabilities of the Scales

	Ability Name	Mission Attack	Mission Cargo	Mission Observ.	Mission Utility
1.	Written Expression	.07	.08	.11	.06
2.	Written Comprehension	.08	.12	.51	.13
3.	Oral Expression	.20	.25	.65	.66
4.	Oral Comprehension	.07	.50	.63	.51
5.	Perceptual Speed	.48	.72	.60	.78
6.	Visualization	.66	.76	.37	.65
7.	Spatial Orientation	.69	.70	.61	.89
8.	Divided Attention	.67	.82	.63	.77
9.	Selective Attention	.74	.69	.57	.77
10.	Flexibility of Closure	.81	.81	.51	.66
11.	Speed of Closure	.70	.76	.39	.69
12.	Reaction Time	.72	.86	.73	.83
13.	Choice Reaction Time	.68	.82	.67	.71
14.	Multilimb Coordination	.79	.84	.74	.85
15.	Control Precision	.80	.88	.77	.79
16.	Kinesthetic Memory	.71	.86	.64	.81
17.	Rate Control	.80	.83	.77	.84
18.	Arm-Hand Steadiness	.84	.83	.66	.79
19.	Finger Dexterity	.71	.62	.54	.44
20.	Speed of Limb Movement	.55	.75	.76	.78
21.	Wrist-Finger Speed	.53	.67	.54	.84
22.	Memorization	.42	.73	.42	.51
23.	Decision Making	.64	.85	.50	.66
24.	Information Ordering	.20	.57	.11	.40
25.	Category Flexibility	.40	.57	.33	.26
26.	Number Facility	.27	.37	.42	.15
27.	Problem Sensitivity	.29	.78	.37	.73
28.	Deductive Reasoning	.47	.78	.39	.67
29.	Inductive Reasoning	.53	.62	.38	.60
30.	Originality	.53	.69	.23	.56
31.	Fluency of Ideas	.31	.71	.24	.69
32.	Stress Tolerance	.59	.85	.84	.86

emphasized that the higher reliability coefficients, however, were associated with the important abilities identified for test development in Table 3 (e.g., Spatial Orientation, Divided Attention, and Kinesthetic Memory).

Establishment of Testing Procedures and Instruments

A battery of ten tests was developed that measured one or more of the critical abilities. These tests and their format as well as the time parameters and scoring mechanisms are presented in Figure 5. The tests were given in the sequence indicated in the figure. It should be emphasized that a part of the conceptual design of the tests developed in the MTAB was based on a previous study conducted by ARRO staff which provided test specifications for many of the computer-assisted tests (Imhoff & Levine, 1980).

A major part of the developmental time was devoted to preliminary evaluation of the battery's technical and administrative soundness. For example, based on a pretest involving approximately 90 student pilots and 30 civilian participants who took either all or parts of the MTAB (June through September, 1982) the instruments were revised. This included improvements in instructions, scoring and programming.

The first five tests in the battery were computer-assisted. The software was developed for the Apple II computer (48k, 16 sector), and was written in Applesoft Basic. Complete listings of each program were in an earlier report (Myers, Jennings, Schemmer, & Fleishman, 1982). The apparatus for testing included a video monitor with a 12-inch (diagonal) screen located on the top of the Apple's keyboard console. Furthermore, there were ten buttons located in a horizontal row every three inches on a panel in front of the Apple console. These were numbered 0 through 9 from left to right. The buttons faced the person taking the tests. Two dual axis control sticks were also attached on the left and right ends of this panel (Kraft Model KJS-01A). The person taking the tests could adjust the location of his/her seat so that they could comfortably reach the buttons and the two control sticks. The tests on

MISSION TRACK ASSIGNMENT BATTERY (MTAB)

Abilities	Tests	Time	Format	Scoring	Source
1. Memorization	Memory	7 Min.	Apple	Number right	ARRO
2. Multilimb Coordination	Complex Coordination	25 Min.	Apple	Time on target	ARRO
3. Kinesthetic Memory	Kinesthetic Memory	11 Min.	Apple	Number right/reaction time	ARRO
4. Divided Attention (including rate control and choice reaction time)	Time Sharing	28 Min.	Apple	Time on target and reaction time/number right	ARRO
5. Perceptual Speed	Perceptual Speed-Apple	5 Min.	Apple	Time to perceive accurately	ARRO
6. Perceptual Speed	Identical Pictures (P3)	3 Min.	Paper/Pencil	Number right-number wrong	ETS
7. Flexibility of Closure	Hidden Patterns (CF2)	6 Min.	Paper/Pencil	Number right-number wrong	ETS
8. Speed of Closure	Gestalt Completion (CS1)	4 Min.	Paper/Pencil	Number right-number wrong	ETS
9. Spatial Orientation	Card Rotations (S1)	4 Min.	Paper/Pencil	Number right-number wrong	ETS
10. Selective Attention	Dichotic Listening	15 Min.	Tape	Number right	Navy

Estimated Time Limits

A.	Introduction	10 Min.
B.	Apple Tests	80 Min.
C.	Dichotic Listening Test	20 Min.
D.	ETS Tests	25 Min.
E.	Background	10 Min.
		145 Min. (2 Hrs. 25 Min.)

Figure 5. Mission Track Assignment Battery (MTAB).

the Apple were self-administered. Each candidate was told in the introduction that the tests were designed to be taken without instructions from the administrator; therefore, it was important that candidates carefully read all of the instructions presented on the screen before completing each test.

The Memory Test measured the ability to remember information. In this test the subject was presented with a sequence of digits (i.e., 1 to 9) and pushed a button corresponding to the item which occurred two digits previously. The task was presented in two parts. In the first part the digits were presented for two seconds followed by a four second inter-stimulus interval so that eight seconds pass between the offset of a digit and the response to that digit. In the second part, the inter-stimulus interval was two seconds so six seconds pass between the offset of a digit and the response to that digit. The number of correct responses in each part was the index of memory retrieval facility.

The Complex Coordination Test measured the ability to coordinate movements of two or more limbs (for example, two arms, two legs or one leg and one arm) together, such as in moving equipment controls. In the Complex Coordination Test, the subject manipulated two independent hand controls to make continuous corrections on three axes. A display of a vertical and a horizontal row of dots intersecting in the center of the screen was presented, along with two response symbols. One symbol was controlled by a joystick, and the person attempted to counteract its movement and keep it at the point of intersection of the two rows of dots (essentially a two-dimensional compensatory tracking task). At the same time a bar marker moved left to right at the bottom of the screen. The subject used a control stick with his left hand and aligned this marker with the vertical row of dots. The sum of error in terms of distance from the desired location of both symbols on all three axes served as the indicators of the individual's multilimb coordination ability.

The Kinesthetic Memory Test measured the ability to sense the position and movement of the body and limbs (e.g., hands, arms, and legs) without looking. It involved the awareness of limb position as well as the ability to accurately reproduce movements without looking. For example, this ability is required in reaching for a control without fumbling while visually concentrating on another task. Each trial in the Kinesthetic Memory Test was presented in two parts. In the first part, a warning tone was sounded, followed by the presentation of a sequence of four digits. The subject responded by pressing the space bar and activating four buttons, which corresponded to the digits presented, in the same sequence as the digits were presented. Three such presentations occurred. The same digit sequence was used in each trial. In the second part of a trial the person was required to wear opaque goggles, and made four attempts to activate the learned sequence of buttons without visual guidance at the sound of the warning tone. The speed and accuracy with which these blind activation sequences were performed served as indices of the kinesthetic sensitivity of the individual. Also, the candidates were told that the administrator would notify them via a verbal command or a tap on the shoulder (if others were being tested in the same room) when to remove the opaque lenses.

The Time Sharing Test measured the ability to shift attention back and forth between two or more sources of information. It is involved in multiple-task situations which require parallel information processing, rapid intertask switching, and allocation of processing resources according to specified priorities (Damos & Lintern, 1981). The person in this test was required to perform a compensatory tracking task and to react as quickly as possible to high and low tones in a choice reaction time task. To perform the compensatory tracking task the individual anticipated the movement of a bar marker on a visual display, and operated a control stick to counteract the movement and keep the marker aligned with a fixed central point. The person was also instructed to press a key on the Apple console corresponding to the tone as quickly as possible. It was assumed that the person had a fixed processing capacity, and that a stable percentage of that capacity was devoted to the

primary tracking task. A person with a larger overall capacity, then, should have reacted more quickly to the digits in the choice reaction time task (secondary). Reaction time in the secondary task was therefore an index of the subject's time sharing ability.

<u>The Perceptual Speed Test</u> measured the ability to perceive rapidly presented visual stimuli. It involved elements of iconic memory, encoding speed, and recognition. Each trial in the Perceptual Speed Test started with an "X" near the middle of the display monitor and the instruction to the subject to press the space bar when ready to start the trial. When the space bar was pressed, the screen went blank and a double beep warning sounds. Two seconds later, the stimulus string of 4 digits was presented at the location of the "X". Following the brief stimulus presentation, the stimulus field was masked by a small box and the subject was then instructed to reproduce the stimulus sequence using the keyboard. The subject's response sequence was displayed until completed and then the next trial started. The test used a threshold seeking algorithm to measure the stimulus presentation duration at which the subject had an equal probability of correctly perceiving and responding to two stimulus sets in a row or of missing one stimulus set. The resulting duration, then, was that at which the subject had a probability of .707 of correctly identifying all four digits in a stimulus set.

The next four tests were in paper-and-pencil format. Each test measured a particular ability found critical to pilot success in helicopter operations (i.e., Perceptual Speed, Flexibility of Closure, Speed of Closure and Spatial Orientation). The tests, developed by Educational Testing Service (ETS), included <u>Identical Pictures</u>, <u>Hidden Patterns</u>, <u>Gestalt Completion</u> and <u>Card Rotations</u>. Instructions to administer each test were provided in the test booklet. The tests were short and speeded and, therefore, required close supervision by the administrator. The administrator read carefully all of the instructions to each candidate, and made sure he/she understood the test requirements before attempting to complete the items. The tests were given individually or in pairs depending on the availability of pilots.

The last test in the battery, designed by Griffin and Mosko (1978) at the Naval Aerospace Medical Laboratory, Pensacola, Florida, measured Selective Attention (i.e., ability to concentrate on a task and not be distracted by irrelevant information). The stimuli in the Dichotic Listening Test were presented via a dual channel tape player. While wearing a headset the candidate sat at a desk and recorded his/her responses (i.e., numbers) on an answer sheet. The messages consisted of series of letters with digits embedded in the series. A voice command told the person which ear to attend to, and his task was to indicate which digits occurred in the relevant ear by writing them on the answer sheet. There were interfering sounds mixed in. The background noise which consisted of VOTRAX digits played in reverse was added to each channel to increase task difficulty. Number of errors in terms of missed digits (omissions) served as the index of the selectivity of the subject's attentional processes. Six practice trials allowed the administrator to determine whether the candidate understood the requirements of the test. Because of the importance in implementing objective, standardized test procedures, the same process for each candidate was followed. All of the instructions were read to each candidate. Answers to questions were allowed before and during the practice trials. Since questions might distract others taking the test no questions were allowed during the 12 test trials.

PHASE III. ANALYSIS OF RESULTS FROM THE MTAB ADMINISTRATION

The purpose of the third phase of the project was to administer the MTAB to a sample of student pilots who took the test before starting IERW and to determine the psychometric properties of the test battery. The students were either Commissioned Officers or students in the Warrant Officer Development Course. The participants were tested in two rooms. The first room contained the apparatus for the Dichotic Listening Test (DLT). In the DLT two individuals were tested at the same time. Each soldier wore a dual channel headset, faced the table and recorded their answers on paper forms. The test monitor gave them both the directions. At the same table they completed the four ETS written tests (i.e., Identical Pictures, Hidden Patterns, Gestalt Completion, and Card Rotation). Next, the participants went to the second room which had two Apple testing consoles. Each participant was seated at the Apple console which included the panel with ten buttons numbered 0 through 9 and the two control sticks located on the left and right ends of the panel. The testing stations were separated by sound resistant partition. The participants could not see each other complete the test.

Description of Sample

The total sample of 275 persons consisted of 140 Warrant Officer Candidates and 135 Commissioned Officers. A more complete description of the sample, based on biographical information is shown in Table 5. To obtain estimates of test-retest reliability a sub-sample of 53 soldiers was tested twice with a two week interval between administrations.

The results from the survey of pilot attitudes and interests are presented in Table 6. It was found that the most often preferred missions were Attack, Utility and Aeroscout, while the least frequently desired mission assignment was Cargo. The primary reason the students decided to participate in IERW was the "desire to fly." Generally, the student pilots felt neutral about the testing process experienced, except they did find it clear and interesting.

TABLE 5

Background Information

Background Items	Total Sample	Test-Retest Subsample (N=52)
Number Who Used		
ARRO Computer	132	24
ARI Computer	143	29
Age (Average Years)	24.6 (.20)	25.5 (.55)
Gender		
Male	260	49
Female	15	4
Race		
White	250	47
Black	14	2
Native American	3	1
Hispanic	3	1
Oriental	3	2
Other	2	0
Education (Average Years)	14.7 (.11)	14.4 (.26)
Entry Source		
Warrant Officer		
In Service	96	21
Civilian Entry	44	7
Total	140	28
Commissioned Officers		
ROTC	84	18
West Point	40	7
OCS	11	0
Total	135	25
Average Number of Flight Hours Before IERW	51.1 (13.1)	25.5 (7.8)
Length of Prior Military Service (Months)	31.4 (1.8)	38.3 (4.8)

Note: Numbers in parenthesis are standard error of the mean.

TABLE 6

Survey of Pilot Attitudes and Interest

Survey Items	Scores
Preferences for Being Trained in a Mission	
Attack	97 (35%)
Utility	96 (35%)
Aeroscout	63 (23%)
Cargo	19 (7%)
Average Time Spent Playing Electronic Games (e.g., Atari)	3.8 hours/month (.60)
Average Typing Proficiency	24.5 words/minute (1.09)
Average Keypunching Proficiency	8.9 words/minute (.87)
Feelings About Testing Process (Mean) (Scale 1 to 10)	
Nervous - Restful	5.0 (.13)
Clear - Confusing	2.3 (.14)
Complex - Simple	4.7 (.13)
Boring - Interesting	6.6 (.14)
Relaxed - Tense	4.3 (.14)
Number Who Attended Vocational School	68 (25%)
The Types of Vocational School	
Construction	7
Heavy Equipment Operator	1
Shop Traders School	6
Electronics School	9
Automobile Mechanic Training	11
Other	35
Sources of Influence to Enlist in IERW Program	
Desire to Fly	245 (89.1%)
Status of Being a Pilot	10 (4%)
Desire for Officer Status	9 (3%)
Encouraged by Superiors	8 (3%)
Salary Level Offered	2 (<1%)
Encouraged by Peers	1 (<1%)

Comparison of Test Performance Across Trials

Descriptive statistics were calculated for all tests that involved repeated testing across more than one trial. The results are presented in Table 7. The Complex Coordination Test included three practice trials involving the one dimensional tracking and three practice trials involving the two dimensional tracking which were followed by fifteen combined trials. The Time Sharing Test consisted of 3 practice trials for the tracking task and 2 practice trials for the reaction time task, which were followed by 10 combined trials. The tracking scores were reported in terms of root mean squared error. Each whole number increase in a mean score represented an increase of .75 mm distance from the target. For example, in the two dimension scores, a score of 35 equaled about 1 inch from target, and in the one dimension scores a score of 30 equaled about 1 inch from target. The screen was 280 dots wide by 190 dots tall.

The findings indicated that for the Complex Coordination, Kinesthetic Memory and the Time Sharing tests, performance improved significantly over trials. The ANOVA indicated that the linear components of the trial effects were significant. In the Complex Coordination Test the two dimension and one dimension tracking performance improved over the combined trials, $F(1,278) = 273$, $p < .001$. The Kinesthetic Memory had significant improvement, $F(1,275) = 6.21$, $p < .01$. In the Time Sharing test tracking performance and number correct in the choice reaction time task improved during the combined test trials, $F(1,276) = 14.57$, $p < .001$ and $F(1,276) = 14.65$, $p < .001$. Thus, practice effect was an important source of variance in test performance. It should be noted that, as expected, there was a significant decrease in tracking performance for the dual task trials, when compared with the single trials.

The variability in each test did not change as a function of practice. There were, however, some differences in the nature of the frequency distributions as practice continued. Although, most test scores were normally distributed, there were a few distributions that were

TABLE 7

Comparison of Performance for Computer-Assisted Tests Across Trials

Test(s)	TRIALS 1	2	3	4	5	6	7	8	9	10
Dichotic Listening Test	8.3[1] (.76)	8.6 (1.0)	8.4 (1.3)	8.8 (.71)	8.9 (.46)	8.6 (.96)	8.6 (.92)	8.5 (.86)	8.8 (.63)	8.7 (.84)
Memory Test	19.5 (4.7)	15.6 (5.2)								
Complex Coordination										
One Dimensional	40.3 (32.5)	22.2 (24.5)	19.6 (20.1)				71.3 (19.5)	50.7 (22.1)	61.8 (19.9)	52.5 (23.7)
Two Dimensional				47.9 (24.4)	43.0 (21.5)	29.0 (19.3)	66.1 (24.7)	62.0 (27.0)	79.5 (26.7)	55.8 (26.7)
Kinesthetic Memory (Number Correct)	3.0 (1.2)	3.3 (1.0)	3.2 (1.1)	3.2 (1.1)	2.9 (1.1)	3.0 (1.1)	3.0 (1.0)	3.0 (1.0)	3.5 (.87)	3.5 (.92)
Time Sharing										
One Dimensional	29.8 (14.0)	28.3 (12.5)	30.8 (14.1)			42.9 (18.2)	41.0 (17.9)	41.4 (18.7)	43.8 (18.7)	40.6 (19.7)
Reaction Time (Number Correct)				46.1 (5.8)	47.7 (4.2)	46.9 (5.1)	47.2 (4.2)	46.7 (5.6)	46.3 (5.3)	46.1 (6.2)
Reaction Time (M. Seconds)				.47 (.08)	.43 (.08)	.53 (.10)	.55 (.09)	.54 (.09)	.56 (.01)	.57 (.10)
Perceptual Speed										
Last 15 trials (Duration[2])	57.6 (36.2)	55.3 (38.0)	50.4 (37.6)	49.0 (37.3)	46.8 (37.2)	45.7 (37.0)	44.4 (36.6)	43.9 (37.0)	42.1 (37.0)	40.9 (36.5)

[1] Combination of Parts A and B.

[2] Scores can be converted to mili seconds by using msecs = 1.64 (stimulus duration) + 124.

Note: The numbers in the table represent means and standard deviations (X).

TABLE 7 (Continued)

Test(s)	TRIALS 11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Dichotic Listening Test	7.7 (1.2)	8.8 (.80)													
Memory Test															
Complex Coordination															
One Dimensional	66.1 (25.2)	51.9 (21.7)	69.4 (16.3)	57.9 (25.9)	49.4 (24.4)	41.7 (23.0)	57.0 (22.1)	45.7 (24.1)	56.2 (27.8)	43.6 (25.8)	40.7 (24.8)				
Two Dimensional	52.5 (26.9)	57.9 (23.7)	60.2 (28.9)	52.6 (24.9)	57.7 (27.8)	73.3 (24.2)	69.6 (24.3)	60.6 (25.5)	65.5 (26.9)	61.9 (24.6)	61.1 (28.0)				
Kinesthetic Memory (Number Correct)	3.4 (1.1)	3.3 (1.1)	3.1 (1.1)	3.1 (1.1)	3.3 (1.0)	3.3 (1.0)	3.2 (.96)	3.1 (1.0))	3.2 (.97)	3.2 (.97)					
Time Sharing															
One Dimensional	40.5 (19.7)	41.4 (19.7)	43.9 (20.2)	37.8 (19.0)	38.9 (19.1)										
Reaction Time (Number Correct)	46.0 (7.2)	46.1 (6.0)	45.8 (6.4)	46.0 (6.4)	45.6 (7.0)										
Reaction Time (M. Seconds)	.56 (.10)	.56 (.10)	.57 (.10)	.56 (.10)	.56 (.10)										
Perceptual Speed															
Last 15 trials (Duration[2])	42.9 (35.8)	42.8 (35.8)	43.0 (35.2)	42.4 (35.0)	39.3 (35.1)										

[1] Combination of Parts A and B.

[2] Scores can be converted to mili seconds by using msecs = 1.64 (stimulus duration) + 124.

Note: The numbers in the table represent means and standard deviations (X).

skewed. For example, in the Complex Coordination Test the score distribution for the two dimension tracking was slightly skewed, but became more normal as practice was allowed. Similarly, the Perceptual Speed and Kinesthetic Memory tests produced scores that were not normally distributed. The degree of deviation, however, decreased over test trials. The Dichotic Listening Test produced non-normal distributions which were high in kurtosis and negatively skewed across all trials. This finding indicated that the test was relatively easy, and thus many student pilots scored at the upper end of the distribution.

A major finding was that most tests had high internal consistency. The internal consistencies, which were used to estimate the reliability of each test, indicated that performance was highly correlated across trials. These results are presented in Table 8. The internally consistent tests indicated that a total of the trial scores could be used as a summary measure for each test in order to provide a reliable measure of that ability. In some cases, such as the Time Sharing and Complex Coordination tests, there was more than one summary measure. For the other tests, such as Memory, Kinesthetic Memory, Dichotic Listening, and the four ETS tests had one summary score. The subsequent data analyses were based on these overall total scores obtained for each test in the battery.

Group Comparisons

The summary scores, presented in Table 9, indicated that performance in all tests was stable when comparing between different computers, ranks, gender and racial groups. It was found that the operating differences between the two models of Apple II computers did not differentially affect test performance. This finding is important, in that it suggested that the two computers were reliable and should yield similar test scores regardless of the machine used. With regard to gender differences the only significant difference was in the Complex Coordination Test where women had significantly lower two-dimensional tracking scores than men. In considering racial differences, the only significant differences were in the Memory, Complex Coordination and

TABLE 8

Internal Consistency Reliability Estimates

Tests	Estimates
Memory	.65*
Complex Coordination	
One Dimensional	.86
Two Dimensional	.94
Kinesthetic Memory	
Number Correct	.82
Response Time	.84'
Time Sharing	
One Dimensional	.97
Choice Reaction Time	.98
Choice Reaction Number Correct	.91
Perceptual Speed (Duration)	
Last 10 Trials	.99
Last 15 Trials	.99
Dichotic Listening Test	.61

* ANOVA and intraclass correlations were used in all tests except Memory which used KR_8 reliability of the sum.

TABLE 9

Descriptive Statistics for Summed Test Scores

Test	Total Sample	Rank: Warrant Officer Candidate (N=140)	Rank: Commissioned Officer (N=135)	Gender: Male (N=260)	Gender: Female (N=15)	Computers: ARI (N=143)	Computers: ARRO (N=132)	Race: White (N=261)	Race: Black (N=14)
Memory	35.1 (8.5)	34.2 (8.4)	35.9 (8.5)	35.1 (8.5)	34.7 (8.9)	34.9 (8.3)	35.3 (8.7)	35.6 (8.2)	28.0* (10.3)
Complex Coordination									
One Dimensional	815.8 (200.0)	829.2 (196.2)	803.3 (203.1)	810.7 (197.1)	910.4 (229.2)	810.5 (202.3)	821.3 (198.1)	813.5 (200.3)	838.4 (220.9)
Two Dimensional	936.2 (285.1)	907.5 (271.4)	958.6 (271.2)	919.1 (273.5)	1178.9* (327.2)	954.8 (294.1)	913.5 (270.3)	925.6 (277.4)	1089.9* (378.1)
Kinesthetic Memory (Blind)									
Number Right	63.9 9.8)	63.4 (9.4)	64.5 (10.2)	64.1 (9.5)	61.5 (15.1)	63.5 (10.6)	64.4 (9.1)	63.7 (9.9)	64.5 (9.0)
Response Time	191.9 (42.0)	198.7 (42.4)	187.1 (41.8)	193.3 (42.8)	166.7 (24.2)	189.6 (43.6)	193.4 (41.5)	188.9 (40.7)	229.1 (56.1)
Time Sharing									
One Dimensional	412.3 (169.8)	405.1 (172.1)	417.7 (168.2)	404.8 (168.6)	527.1 (156.4)	412.2 (177.6)	410.8 (163.3)	412.9 (172.2)	399.3 (175.8)
Reaction Time/Number Right	462.4 (45.7)	461.4 (49.7)	464.3 (39.3)	462.7 (44.3)	466.8 (52.6)	459.9 (52.6)	465.6 (35.9)	461.7 (46.3)	476.9 (10.5)
Reaction Time	5.6 (.9)	5.7 (.9)	5.5 (.88)	5.5 (.9)	5.9 (.7)	5.5 (.98)	5.6 (.81)	5.5 (.8)	5.7 (1.1)
Perceptual Speed									
Duration	476.6 (367.1)	489.2 (377.6)	460.7 (358.1)	480.6 (373.1)	372.5 (235.6)	490.8 (373.9)	459.6 (361.8)	676.0 (544.2)	798.2 (464.6)
Response Time	38.9 (7.1)	39.3 (7.4)	38.4 (7.0)	39.0 (6.9)	35.9 (9.7)	38.6 (7.3)	39.0 (7.1)	38.6 (7.0)	40.1 (7.4)
Identical Pictures	72.3 (13.8)	70.1 (14.2)	74.4 (13.3)	72.1 (13.9)	76.8 (12.2)	71.8 (14.2)	72.8 (13.6)	72.6 (13.5)	66.6 (17.1)
Hidden Patterns	216.9 (41.9)	214.5 (42.5)	219.3 (41.1)	215.7 (41.0)	237.8 (51.4)	214.1 (40.9)	219.6 (42.6)	217.5 (41.3)	206.4 (39.7)
Gestalt Completion	12.3 (3.3)	12.0 (3.4)	12.5 (3.3)	12.3 (3.4)	11.5 (2.3)	12.6 (3.4)	12.0 (3.3)	12.3 (3.3)	11.4 (4.3)
Card Rotations	121.8 (21.4)	120.5 (21.3)	123.1 (21.6)	122.7 (20.8)	107.6 (28.2)	120.4 (21.7)	123.1 (21.2)	122.9 (21.3)	112.6 (24.4)
Dichotic Listening	102.7 (4.7)	102.6 (5.3)	102.7 (4.1)	102.6 (4.7)	103.7 (3.8)	102.8 (4.2)	102.5 (5.1)	102.9 (3.9)	98.9* (10.7)

* $p < .05$

Dichotic Listening tests. The comparisons between race and gender, however, were limited since there were only 15 females and 14 blacks in the sample. Furthermore, the test variances for these subgroups were usually significantly different from each other and the distributions were highly skewed.

Test-Retest Reliability

As mentioned previously, the internal consistency reliability estimates were relatively high. Test-retest (two week interval) reliability coefficients were also calculated (Table 10). Overall, the test-retest coefficients were at the moderate level for the computer-based tests (i.e., average was .52). The Time Sharing Test was the most reliable, especially for the tracking task and the choice reaction time. Performance in the Kinesthetic Memory Test had the lowest test-retest reliability over the two week time period. For the written tests, the four ETS tests had high reliabilities ranging from .72 to .91, while the Dichotic Listening Test yielded a coefficient of .56.

The test-retest analysis did reveal potential areas where the tests could be improved. The reliability for the Time Sharing Test could be improved by eliminating the number right scores in the choice reaction task; therefore, the Time Sharing Test would be scored using only the tracking and reaction time components. Although it may be argued that more costly control sticks might help to improve test reliabilities, the tracking task in the Time Sharing Test had sufficient reliability (.72). However, the lower reliabilities for the Complex Coordination Test may have been due to the dynamic nature of information processing and cognitive abilities. In order for tracking scores to remain consistent in the test-retest analysis, the individuals would have to remember the specific strategy used in the initial testing session. The low reliabilities may have been due to poor recall rather than to a change in motor skills. Also, the results indicated that tracking performance was still improving at the end of the first session. Therefore the low reliabilities may reflect fluctuations in individual differences in

TABLE 10

Test-Retest Reliability Coefficients

Test	Reliabilities
Memory	.60
Complex Coordination	
One Dimensional	.43
Two Dimensional	.58
Kinesthetic Memory	
Number Correct	.44
Response Time	.21
Time Sharing	
One Dimensional	.72
Choice Reaction Time	.67
Choice Reaction Number Correct	.35
Perceptual Speed (Duration)	
Last 10	.58
Last 15	.59
Identical Pictures	.72
Hidden Patterns	.77
Gestalt Completion	.91
Card Rotations	.72
Dichotic Listening	.56

learning. It is still important to note that performance at one point in time was significantly predictive of performance two weeks later. Perhaps a longer scored segment of performance would increase test-retest reliabilities. This possibility could be examined in later research.

Before any major changes in the design of software are made it would be desirable to collect more data. It would be desirable, for example, to collect additional test-retest data using a variety of inter-test time periods and a larger sample.

Factor Analysis

A maximum likelihood factor analysis with a varimax rotation was carried out to identify the commonalities among the different elements making up the MTAB. The analysis yielded five factors (i.e., eigenvalues greater than 1.0) which explained 40 percent of the variance in test performance (Table 11).

Factor 1 appeared to be primarily involved with tracking skills. It accounted for 22 percent of the variance. The tracking scores in the Complex Coordination and the Time Sharing tests loaded highest on the factor. The tasks required the student pilots to use a control stick for counteracting the movement of a cursor and keeping it in alignment with the target in the screen (i.e., compensatory tracking).

Factor 2 accounted for 6 percent of variance and involved tests that required a series of accurate responses by the pilots. The measures which were scored using the number of correct responses in both the Time Sharing and the Perceptual Speed tests loaded highest on the second factor. Other tests such as Memory, Kinesthetic Memory, and Dichotic Listening which tabulated the number of correct responses as dependent scores also loaded on the second factor.

TABLE 11

Factor Analysis of MTAB Tests

Tests	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Communality Estimates
Complex Coordination (One Dimensional)	80	-14	04	-13	-06	67
Complex Coordination (Two Dimensional)	71	-05	10	-12	-18	56
Time Sharing (Tracking)	66	-21	38	-07	-08	62
Time Sharing (Number Right)	06	55	-32	-05	-07	41
Perceptual Speed (Response Time)	10	06	55	-12	-06	32
Identical Pictures	-04	06	-21	75	15	63
Card Rotations	-21	15	-14	09	74	64
Perceptual Speed (Number Right)	08	-45	-04	-14	-10	24
Memory	-12	29	-26	14	24	24
Kinesthetic Memory (Number Right)	-04	31	03	05	12	11
Hidden Patterns	-10	28	-33	36	31	42
Gestalt Completion	-09	05	01	33	-01	12
Time Sharing (Reaction Time)	29	-24	41	-02	-14	33
Dichotic Listening	-07	42	-05	02	01	18
Sum of Squared Loadings	1.76	1.08	.97	.90	.82	

Factor 3, which accounted for 5 percent of the variance, dealt mostly with response time measures in the Perceptual Speed and Time Sharing tests. Factor 4 appeared to involve tests that required the ability to see differences and similarities among different types of pictures. This factor accounted for 3 percent of the variance. Factor 5, accounted for 3 percent of the variance, and although it was somewhat difficult to interpret the test that loaded on the factor involved spatial orientation.

Although the study was not originally designed for factor analysis, these results provided some understanding of the different independent dimensions which represented the performance domain involving both written and computer-based testing protocols. The analysis demonstrated that tasks making up each test tended to load on the same factor. Because of the independence found among the tests there is reason to believe that in a validation study the five domains should not turn out to be redundant, overlapping predictors. The analysis also might provide guidance in future attempts to revise the test battery. For example, separate tests loading on the same factor might be combined into one test.

Correlations Between MTAB Tests and Biographical Data

Correlations between tests and relevant background information obtained during testing were determined. The results, presented in Table 12, demonstrated that only a few correlations reached a level of practical significance. Although age, gender, and level of education generally were not correlated with test performance, age was negatively correlated with the Dichotic Listening Test (-.25 $p < .01$) and gender was associated with performance in the Complex Coordination Test (i.e., two dimensional tracking, +.20 $p < .01$).

The number of flight hours obtained previously to testing and experience playing electronic games were not correlated with test performance. There was, however, a low significant correlation between experience playing electronic games and performance in the Complex

TABLE 12

Correlations Between Tests and Biographical/Attitude Data

	Variables																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1.00																
2	-.00	1.00															
3	-.09	-.07	1.00														
4	-.06	-.04	.10	1.00													
5	-.10	.09	.00	-.08	1.00												
6	.10	.60	-.13	-.25	.08	1.00											
7	.11	-.13	-.03	-.08	-.07	-.02	1.00										
8	.01	.06	-.01	.09	-.01	.08	.00	1.00									
9	.00	.01	.06	.12	.02	.03	.00	.64	1.00								
10	.02	.01	-.05	.11	-.02	.05	-.03	.34	.18	1.00							
11	-.02	-.01	.02	.08	.05	.07	-.04	.29	.40	.82	1.00						
12	.02	.14	-.15	-.02	.04	.03	-.04	-.08	-.03	.06	.04	1.00					
13	.01	-.01	-.01	.00	-.05	.01	.06	-.03	-.09	-.06	-.10	-.12	1.00				
14	-.07	.03	-.08	-.00	.01	-.03	-.08	-.02	-.06	.08	.04	.35	-.03	1.00			
15	-.01	.10	-.02	-.14	.06	.10	.11	.04	.08	.08	.14	.09	-.13	-.10	1.00		
16	.08	-.14	.12	.05	-.05	-.00	.11	.08	.11	-.01	.07	-.63	.17	-.39	-.06	1.00	
17	-.02	.16	-.03	-.31	.11	.16	.05	-.03	-.08	-.10	-.09	.00	.07	.05	.07	-.07	1.00

LEGEND

1. COMPUTER
2. AGE
3. SEX
4. EDUCATION
5. FLIGHT HOURS
6. PRIOR VOCATIONAL TRAINING
7. COMPUTER GAMES
8. TYPE (?)
9. TYPE (WPM)
10. KEYPUNCH (?)
11. KEYPUNCH (WPM)
12. NERVOUS/RESTFUL
13. CLEAR/CONFUSED
14. COMPLEX/SIMPLE
15. BORING/INTERESTING
16. RELAXED/TENSE
17. VOCATIONAL TRAINING (?)
18. MEMORY
19. COMPLEX COORDINATION (TWO DIMENSIONAL)
20. COMPLEX COORDINATION (ONE DIMENSIONAL)
21. KINESTHETIC MEMORY (NUMBER RIGHT)
22. KINESTHETIC MEMORY (RESPONSE TIME)
23. TIME SHARING (TRACKING)
24. TIME SHARING (NUMBER RIGHT)
25. TIME SHARING (REACTION TIME)
26. PERCEPTUAL SPEED (10 DURATION)
27. PERCEPTUAL SPEED (15 DURATION)
28. PERCEPTUAL SPEED (10 RESPONSE TIME)
29. PERCEPTUAL SPEED (15 RESPONSE TIME)
30. IDENTICAL PICTURES
31. HIDDEN PATTERNS
32. GESTALT COMPLETION
33. CARD ROTATION
34. DICHOTIC LISTENING

TABLE 12 (Continued)

	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
18	1.00																
19	-.22	1.00															
20	-.13	.58	1.00														
21	.15	-.04	-.08	1.00													
22	-.34	.12	.07	-.13	1.00												
23	-.30	.53	.56	-.12	.07	1.00											
24	.12	-.18	-.14	.15	-.07	-.37	1.00										
25	.00	.00	.00	.00	.00	.00	.00	0.00									
26	-.20	.08	.18	-.14	.28	.16	-.20	0.00	1.00								
27	-.20	.08	.18	-.15	.28	.16	-.20	0.00	1.00	1.00							
28	-.24	.15	.13	.02	.23	.23	-.15	0.00	-.03	-.02	1.00						
29	-.26	.14	.14	.05	.22	.24	-.13	0.00	-.04	-.03	.95	1.00					
30	.23	-.18	-.18	.07	-.16	-.19	.02	0.00	-.14	-.14	-.22	-.24	1.00				
31	.31	-.24	-.23	.10	-.16	-.30	.27	0.00	-.19	-.19	-.22	-.22	.42	1.00			
32	.09	-.16	-.11	.02	.03	-.11	-.02	0.00	-.06	-.05	.03	-.02	.24	.17	1.00		
33	.32	-.30	-.25	.10	-.32	-.28	.09	0.00	-.15	-.16	-.14	-.15	.25	.38	.04	1.00	
34	.18	-.13	-.15	.05	-.12	-.11	.14	0.00	-.17	-.17	-.04	-.04	.05	.12	.14	.10	1.00

LEGEND

1. COMPUTER
2. AGE
3. SEX
4. EDUCATION
5. FLIGHT HOURS
6. PRIOR VOCATIONAL TRAINING
7. COMPUTER GAMES
8. TYPE (?)
9. TYPE (WPM)
10. KEYPUNCH (?)
11. KEYPUNCH (WPM)
12. NERVOUS/RESTFUL
13. CLEAR/CONFUSED
14. COMPLEX/SIMPLE
15. BORING/INTERESTING
16. RELAXED/TENSE
17. VOCATIONAL TRAINING (?)
18. MEMORY
19. COMPLEX COORDINATION (TWO DIMENSIONAL)
20. COMPLEX COORDINATION (ONE DIMENSIONAL)
21. KINESTHETIC MEMORY (NUMBER RIGHT)
22. KINESTHETIC MEMORY (RESPONSE TIME)
23. TIME SHARING (TRACKING)
24. TIME SHARING (NUMBER RIGHT)
25. TIME SHARING (REACTION TIME)
26. PERCEPTUAL SPEED (10 DURATION)
27. PERCEPTUAL SPEED (15 DURATION)
28. PERCEPTUAL SPEED (10 RESPONSE TIME)
29. PERCEPTUAL SPEED (15 RESPONSE TIME)
30. IDENTICAL PICTURES
31. HIDDEN PATTERNS
32. GESTALT COMPLETION
33. CARD ROTATION
34. DICHOTIC LISTENING

TABLE 12 (Continued)

	Variables																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
18	.03	-.14	-.02	.13	-.11	-.17	.03	.10	.19	.02	.10	.07	-.12	-.01	-.01	.03	-.18
19	-.07	.04	.21	.09	-.04	.08	-.21	-.05	-.13	.03	-.03	-.06	-.08	-.01	-.19	-.01	-.08
20	.03	.15	.11	-.02	.04	.15	-.18	.04	-.01	.03	.00	-.05	-.08	-.06	-.17	.04	.09
21	.02	.01	.03	.05	.03	-.03	-.06	-.01	.01	-.00	.04	.08	-.10	.08	-.02	-.01	-.01
22	.04	.11	-.15	-.12	.05	.11	-.06	.06	-.12	.09	.09	.10	.13	.08	.03	.08	.05
23	-.00	.02	.16	.00	-.07	.00	-.16	-.03	-.08	-.00	-.06	-.04	.04	-.02	-.23	.02	.10
24	.03	.09	.01	.03	-.00	.04	-.04	.13	.12	.08	.10	.08	.07	.01	.19	-.01	-.07
25	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
26	-.04	-.01	-.07	-.05	-.07	.02	-.11	-.02	-.02	.06	.09	.04	-.04	.03	-.09	-.00	.02
27	.04	-.01	-.06	-.05	-.06	.02	-.10	-.01	-.02	.06	.09	.04	-.04	.03	-.09	-.01	.02
28	.02	.12	-.10	-.07	.10	.05	.05	-.13	-.23	-.08	-.19	.03	-.05	-.00	.06	-.07	.08
29	.01	.11	-.11	-.09	.09	.04	.05	-.15	-.24	-.10	-.21	.02	-.01	.02	.03	-.05	.13
30	.03	-.24	.08	.09	-.11	-.20	.05	.04	.08	-.00	.01	-.06	-.13	.07	-.01	.08	-.14
31	.07	-.17	.12	.04	-.01	-.15	-.03	.05	.12	-.04	.02	-.01	.04	-.06	.05	.03	-.09
32	-.08	-.12	-.06	-.01	-.01	-.06	.06	.06	.12	.03	.02	-.04	-.09	.07	.06	.02	-.09
33	.06	-.15	-.16	.02	-.06	-.13	.06	.01	.08	-.07	.00	.05	-.01	-.02	-.05	-.12	-.14
34	-.03	-.25	.05	.03	-.04	-.19	.01	-.05	.08	-.00	.02	.00	-.05	-.09	.20	.07	-.06

LEGEND

1. COMPUTER
2. AGE
3. SEX
4. EDUCATION
5. FLIGHT HOURS
6. PRIOR VOCATIONAL TRAINING
7. COMPUTER GAMES
8. TYPE (?)
9. TYPE (WPM)
10. KEYPUNCH (?)
11. KEYPUNCH (WPM)
12. NERVOUS/RESTFUL
13. CLEAR/CONFUSED
14. COMPLEX/SIMPLE
15. BORING/INTERESTING
16. RELAXED/TENSE
17. VOCATIONAL TRAINING (?)
18. MEMORY
19. COMPLEX COORDINATION (TWO DIMENSIONAL)
20. COMPLEX COORDINATION (ONE DIMENSIONAL)
21. KINESTHETIC MEMORY (NUMBER RIGHT)
22. KINESTHETIC MEMORY (RESPONSE TIME)
23. TIME SHARING (TRACKING)
24. TIME SHARING (NUMBER RIGHT)
25. TIME SHARING (REACTION TIME)
26. PERCEPTUAL SPEED (10 DURATION)
27. PERCEPTUAL SPEED (15 DURATION)
28. PERCEPTUAL SPEED (10 RESPONSE TIME)
29. PERCEPTUAL SPEED (15 RESPONSE TIME)
30. IDENTICAL PICTURES
31. HIDDEN PATTERNS
32. GESTALT COMPLETION
33. CARD ROTATION
34. DICHOTIC LISTENING

Coordination Test (i.e., two dimensional). Only the first tracking test given was influenced by prior experience with video games; therefore, it appeared that when the student pilots had had a chance to become familiar with the control apparatus this influence was eliminated.

Generally, prior experiences with keyboards like the Apple were not related to test performance. However, self-reported typing and keypunching speed (i.e., WPM) were significantly correlated with the measures of response time in the Perceptual Speed Test (i.e., -.24 and -.20 $p < .01$, respectively). Finally, the emotional reactions to the testing process experienced by the student pilots were generally uncorrelated with performance in each test in the battery. An exception was the Time Sharing Test, where the "more interested" students did better on the tracking task.

CONCLUSIONS AND SUGGESTED FUTURE RESEARCH

The test battery developed represents a broad range of abilities and skills that were judged by experienced pilots to be critical to successful and safe helicopter operations. The Mission Track Assignment Battery (MTAB) was developed on the basis of a taxonomic approach to job analysis that linked piloting tasks with ability requirements. The present research effort involved extensive programming and pre-testing of the entire battery. Finally, the MTAB was the focus of a large scale evaluation and analysis which determined the psychometric characteristics for each test. The products of that effort were:

- Identification of the tasks considered critical to piloting helicopters in four different mission tracks (i.e., Attack, Aeroscout, Utility, and Cargo);
- Identification of the skills and abilities required to perform these missions;
- Establishment of critical tasks that can be used as a basis for criterion development;
- Development of a battery of tests (called the MTAB) designed to measure each of the skills and abilities identified;
- Development of a computer-interactive testing system to include in this battery;
- Development of scoring procedures and data analytic system for the test battery;
- Development of software documentation and test administration manuals for the test battery; and
- Documentation of the psychometric properties of the different components in the battery.

The findings indicated that most of the tests had high internal consistency reliabilities; individual differences in test performance were consistent across immediately successive trials. Test-retest reliabilities during a two week period were lower, but significant predictions across this interval were achieved. Except for the Kinesthetic Memory Test, the reliabilties for the computer tests were useful for battery inclusion. The written tests developed by ETS (i.e.,

Identical Pictures, Card Rotations, Gestalt Completion, and Hidden Patterns) and the computer-based test measuring time sharing did yield higher test-retest reliabilities (i.e., tracking task and choice reaction time). Although, there are ways to enhance test reliability (e.g., more standardized test administation procedures), it is desirable to collect additional data using a variety of inter-test time periods and a larger sample.

Performance in tests which involved tracking tasks did show significant improvement across trials. Therefore in validating the MTAB it might be useful to use rate of learning scores as one method of scoring the predictors. There are a variety of different procedures which could be used to derive such scores. It may be that students whose tracking scores improve the most would have the greatest probability of success in mission training.

The Dichotic Listening Test (DLT) was found to be an easy test in that most student pilots scored in the upper end of the distribution. Additional research is needed to redesign the test, perhaps making it more difficult by using different background noise. It may be possible to use a different version of the sound track which included VOTRAX digits repeated in reverse order. Although other noise sources might be considered, a possible approach to increasing test difficultly would be to combine the DLT with other computer-based tasks. For example, it might be possible to design a test which includes three components--single tracking task, choice reaction time, and the DLT. This might require the use of all limbs. The difficulty level associated with such a test might better reflect the complex demands of piloting a helicopter in a combat mission.

The findings provided insight into the factor structure of the MTAB. The analysis yielded five factors indicating that in validation those dimensions should not represent overlapping, redundant predictors. The amount of variance accounted for might be increased by adding more tests to the battery. The results from the abilities analysis obtained in the present study would provide the basis for future test development efforts (e.g., Rate Control and Oral Comprehension).

Additional test trials appeared to reduce the impact of prior keyboard experience (e.g., games). Therefore, it might be desirable to include a formal practice session for the computer-based tests. The session should allow the participant to warm-up and become familiar with the keyboard and the control sticks.

Finally, most tests scores were not significantly different when comparing racial and gender groups. Because these subsamples were small, there is a need for additional research which increases the representativeness of women and minorities. In contrast, the research also found that test performance did not vary as a function of the specific Apple computer used nor with the rank of the individual (i.e., WOC and CO).

The present study represents a significant step forward in the use of computer-interactive tests in the measurement of abilities and skills in the perceptual-motor domain. A number of problems in designing and administering computer-interactive test batteries were solved, and a battery was developed that meets the requirements of "content validity". However, there is a need to do criterion related validation research before the battery is used to select and assign pilots in an operational setting.

The validity and utility of the MTAB should be evaluated in terms of its relationship with actual criterion measures which represent pilot success and proficiency in training and in the field. Although there exist several criterion measures currently used in training (e.g., academic and flight grades), there is a need to improve the reliability of these types of training criteria as well as to develop new and improved performance-based measures. For example, the task analysis information obtained in the present effort can be used to develop relevant task-based measures which represent different aspects of pilot proficiency. The total set of measures could then be administered by instructor pilots (SIPs) during checkrides in training and after assignment to a field unit. The data generated from the measures would be used to validate the MTAB as well as to evaluate the progress being made by pilots who are in training or in annual standardized checkrides.

In addition to the validation there is a need to determine the utility of the MTAB in terms of its cost-effectiveness. Estimates of the costs associated with attrition in training or in damaged equipment should be made and then statistically combined with the validity coefficient. The results would demonstrate the cost-benefits of using the MTAB in an operational setting.

Finally, the computer testing technology developed in the present study needs to be evaluated for other Army jobs (e.g., tank operator and maintenance). Presently there have been some preliminary and fragmentary efforts by the Army to adapt computer-interactive systems designed to measure psychomotor abilities. However, these efforts currently need to be coordinated and based on a common framework solidly grounded by generic concepts and methods derived from basic research. The products of this research would be a prototype computer-interactive test battery of tests of generic psychomotor abilities applicable across a wide variety of Army jobs.

REFERENCES

Crosby, J. V. Cognitive pre-training: An aid in the transition from instrument to composite flying. (AFHRL-TR-77-62). Williams AFB, Arizona: Air Force Human Resources Laboratory, October 1977.

Damos, D. L. Residual attention as a predictor of pilot performance. Human Factors, 1978, 20(4), 435-440.

Damos, D. L., & Lintern, G. A comparison of single-and dual-task measures to predict simulator performance of beginning student pilots. Ergonomics, 1981, 24(9), 673-684.

Dohme, J. A. Assignment of Army aviator trainees to undergraduate aeroscout mission training. Ft. Rucker, Alabama: Army Research Institute Field Unit, October 1979.

Farrell, J. P., & Fineberg, M. L. Specialized training versus experience in helicopter navigation of extremely low altitude. Human Factors, 1976, 18(3), 305-308.

Fleishman, E. A. Evaluation of psychomotor tests for the prediction of pilot success: Direction control and compensatory balance tests (Technical Report AFPTRC-TR-54-131). Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center, December 1954.

Fleishman, E. A. Psychomotor selection tests: Research and application in the United States Air Force. Personnel Psychology, 1956, 9, 449-467.

Fleishman, E. A. The structure and measurement of physical fitness. Englewood Cliffs: Prentice-Hall, 1964.

Fleishman, E. A. Toward a taxonomy of human performance. American Psychologist, 1975, 30(2), 1127-1130.

Fleishman, E. A. Systems for describing human tasks. American Psychologist, 1982, 37(7), 821-834.

Fleishman, E. A., & Hogan, J. C. A taxonomic method for assessing the physical requirements of jobs: The physical abilities analysis approach (Final Report). Washington, D.C.: Advanced Research Resources Organization, 1978.

Fleishman, E. A., & Ornstein, G. N. An analysis of pilot flying performance in terms of component abilities. Journal of Applied Psychology, 1960, 44, 146-155.

Fleishman, E. A., Quaintance, M. K., & Broedling, L. A. Taxonomies of human performance: The description of human tasks. New York, NY: Academic Press, 1983.

Gerlach, V. S. Cues, feedback, and transfer in undergraduate pilot training. Phase III. Tempe, Arizona: Arizona State University Department of Education Technology, October 1974.

Gopher, D., & Kahneman, D. Individual differences in attention and the prediction of flight criterion. Perceptual and Motor Skills, 1971, 33, 1335-1342.

Griffin, G.R. & Mosko, J.D. Initial development and evaluation of a test of selective attending (dichotic listening) for the Naval Aviation Training Program: A progress report. Naval Aerospace Medical Research Laboratory, Naval Air Station, Pensacola, Florida, 1978.

Hunter, D. R. Development of an enlisted psychomotor/perceptual test battery. (AFHRL-TR-75-60). Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, Personnel Research Division, November 1975.

Imhoff, D. L., & Levine, J. M. Development of a perceptual-motor and cognitive performance task battery for pilot selection (Final Report). Washington, D.C.: Advanced Research Resources Organization, 1980.

Jensen, R. S., & Benel, R. A. Judgment evaluation and instruction in civil pilot training. Illinios University at Urbana-Champaign: Savoy Aviation Research Laboratory, December 1977.

Levine, J. M., Ogden, G. D., Eisner, E. J. Measurement of workload by secondary tasks: A review and annotated bibliography. Washington, D. C.: Advanced Research Resources Organization, 1978.

Locke, E. A., Zavala, A., & Fleishman, E. A. Studies of helicopter pilot performance: Phase II, The analysis of task dimensions. Human Factors, 1965, 7, 285-302.

McGrevy, D. F., & Valentine, L. D., Jr. Validation of two aircrew psycho-motor tests (AFJR-TR-74-4). Lackland Air Force Base, TX: Air Force Human Resources Laboratory, Personnel Research Division, 1974.

McLaurin, W. A. Validation of a battery of performance tests for prediction of aerospace group equipment course grades (AFHRL-TR-73-20). Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, November 1973.

Melton, A. E. (Ed.), Apparatus tests. (Air Force Aviation Psychology Program Research Report No. 4). Washington, D.C.: U.S. Government Printing Office, 1947.

Miller, J. T., Eschenbrenner, J., Marco, R. A., & Dohme, J. A. Mission track selection process for the Army Initial Rotary Wing Flight Training Program. St. Louis, MO: McDonnell Douglas Astronautics Co., 1981.

Myers, D. C., Gebhardt, D. L., & Fleishman, E. A. Development of physical ability performance standards for Army jobs (Final Report). Washington, D.C.: Advanced Research Resources Organization, 1979.

Myers, D. C., Jennings, M. C., Fleishman, E. A. Validation of cognitive and perceptual-motor tests for assigning pilots to different missions: Job analysis findings. Washington, D.C.: Advanced Research Resources Organization, 1982.

Myers, D. C., Jennings, M. C., Schemmer, M., & Fleishman, E. A. Validation of cognitive and perceptual-motor tests for assigning pilots to different missions: Development and administration of computer-interactive tests. Washington, D.C.: Advanced Research Resources Organization, 1982.

North, R. A., & Gopher, D. Measures of attention as predictors of flight performance. Human Factors, 1976, 18, 1-14.

Pew, R. W., & Adams, M. J. Development of procedures for assignment of subjects in ASUPT experiments: Background and data collection plan (Report No. 3204). Cambridge, MA: Bolt, Beranek, and Newman, Inc., November 1975.

Shrout, P. E., & Fleiss. Intra-class correlations: User assessment of rater reliability. Psychological Bulletin, 1979, 86, 420-428.

Stone, L. W., Krueger, G. P., Holt, W. R. Pursuit rotor tracking performance in conjuction with extended flight operations in a helicopter simulator. Biomedical Applications Research Division, U.S. Aeromedical Research Laboratory, Fort Rucker, Alabama, August 1982.

Theologus, G. C., Romashko, T., & Fleishman, E. A. Development of a taxonomy of human performance: A feasibility study ability dimensions for classifying human tasks. JSAS Catalog of Selected Documents in Psychology, 1973, 3, 15-26.

Winer, B. J. Statistical principles of experimental design. New York, NY: McGraw-Hill Publishing Company, Inc., 1971.

Wierwille, W. W., & Connor, S. A. Evaluation of 20 workload measures using psychomotor tasks in a moving base aircraft simulator. Human Factors, 1983, 25(1), 1-16.

Youngling, E. W., Levine, S. H., Mocharnuk, J. B., & Weston, L. M. Feasibility study to predict combat effectiveness for selected military roles: Fighter pilot effectiveness (Final Report). St. Louis, MO: McDonnell Douglas Astronautics Co., 1977.

Zavala, A., Locke, E. A., Van Cott, H. P., & Fleishman, E. A. Studies of helicopter pilot performance: Phase I, The analysis of maneuvers dimensions. _Human Factors_, 1965, _7_, 273-283.